

Geo-spatial Information Science

ISSN: 1009-5020 (Print) 1993-5153 (Online) Journal homepage: <http://www.tandfonline.com/loi/tgsi20>

Quality assessment of OpenStreetMap data using trajectory mining

Anahid Basiri, Mike Jackson, Pouria Amirian, Amir Pourabdollah, Monika Sester, Adam Winstanley, Terry Moore & Lijuan Zhang

To cite this article: Anahid Basiri, Mike Jackson, Pouria Amirian, Amir Pourabdollah, Monika Sester, Adam Winstanley, Terry Moore & Lijuan Zhang (2016) Quality assessment of OpenStreetMap data using trajectory mining, Geo-spatial Information Science, 19:1, 56-68, DOI: [10.1080/10095020.2016.1151213](https://doi.org/10.1080/10095020.2016.1151213)

To link to this article: <http://dx.doi.org/10.1080/10095020.2016.1151213>



© 2016 Wuhan University. Published by Taylor & Francis Group



Published online: 25 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 1077



View related articles [↗](#)



View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at
<http://www.tandfonline.com/action/journalInformation?journalCode=tgsi20>

Quality assessment of OpenStreetMap data using trajectory mining

Anahid Basiri^a , Mike Jackson^a, Pouria Amirian^b, Amir Pourabdollah^{a,c}, Monika Sester^d, Adam Winstanley^e , Terry Moore^a and Lijuan Zhang^f

^aThe Nottingham Geospatial Institute, The University of Nottingham, Nottingham, UK; ^bOrdnance Survey, Southampton, UK; ^cThe School of Computer Science, The University of Nottingham, Nottingham, UK; ^dInstitute of Cartography and Geoinformatics (IKG), Leibniz Universität, Hannover, Germany; ^eDepartment of Computer Science, Maynooth University, Maynooth, Ireland; ^f Faculty of Civil and Environmental Engineering, The Technion-Israel Institute of Technology, Haifa, Israel

ABSTRACT

OpenStreetMap (OSM) data are widely used but their reliability is still variable. Many contributors to OSM have not been trained in geography or surveying and consequently their contributions, including geometry and attribute data inserts, deletions, and updates, can be inaccurate, incomplete, inconsistent, or vague. There are some mechanisms and applications dedicated to discovering bugs and errors in OSM data. Such systems can remove errors through user-checks and applying predefined rules but they need an extra control process to check the real-world validity of suspected errors and bugs. This paper focuses on finding bugs and errors based on patterns and rules extracted from the tracking data of users. The underlying idea is that certain characteristics of user trajectories are directly linked to the type of feature. Using such rules, some sets of potential bugs and errors can be identified and stored for further investigations.

ARTICLE HISTORY

Received 15 September 2015
Accepted 19 December 2015

KEYWORDS

Spatial data quality;
OpenStreetMap (OSM);
trajectory data mining

1. Introduction

Although OpenStreetMap (OSM) data have been widely used by a range of different applications, its reliability and accuracy have been questioned since most contributors are not doing so as geospatial data experts (Hashemi and Abbaspour 2015; Salk et al. 2015). The quality aspects of OSM have been investigated by different researchers (Amirian et al. 2015; Arsanjani et al. 2015; Fan et al. 2014; Helbich et al. 2012; Koukoletsos, Haklay, and Ellul 2012). A recent study has shown the essentiality of an expert-validation phase for OSM data quality assurance (Salk et al. 2015). Many of current quality-assurance-related applications have focused on comparing OSM data with other sources of data, such as from Google Maps and Ordnance Survey (UK) to evaluate OSM positional, temporal, and thematic accuracy and completeness of coverage.

Data-validation methods, such as node spacing on specified feature polygons in an OSM snapshot, can either be based on user-checking or on predefined automated rules which can detect and correct bugs and errors. Such rules are mainly based on logical assumptions and mapping agencies specification standards, for example, that two crossing roads at the same level must have an intersection.

The idea presented in this paper is to analyze users' travel behaviors to derive rules, which can be used for

checking and validating the quality of user-generated data. This paper focuses on this approach, which is finding bugs and errors based on spatial knowledge extracted from anonymous tracking data of users.

Rules for error detection and data correction, based on detected categories or specific types of error, can be also based on users' travel behaviors and patterns. Thus, it is possible to learn rules and recognize patterns over trajectories and use these to check OSM data validity. Such patterns and rules may highlight anomalies and unusual data patterns as potential errors which can then be manually examined or in some cases automatically corrected. For example, from the analysis of tracking data, some indoor corridors which have been wrongly tagged as "tunnel" can be found. This can be done by analyzing trajectories, categorizing them into car, pedestrian, bicycle, wheelchair, bus, and tram. This analysis is based on each trajectory's average speed and some recognizable patterns such as sequential stops. From the travel mode of the trajectories, the underlying path can be directly categorized. For example, if only groups of pedestrians and wheelchairs use an indoor path (which has been tagged as a road tunnel) and there is no car or bus trajectory matching that path, it is possible to store this path in a "potentially wrongly tagged features" dataset for further quality processing. Figure 1 shows a path, wrongly tagged as "tunnel", that has been calculated as

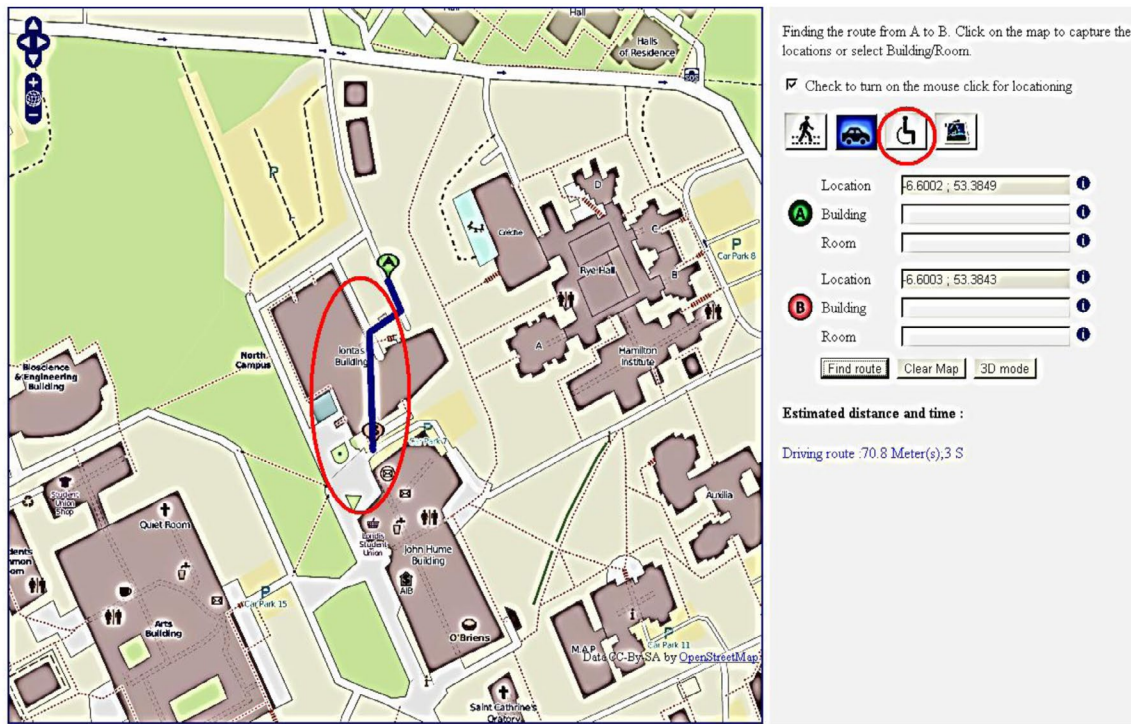


Figure 1. An example of wrongly tagged feature in OSM: an indoor corridor wrongly tagged as a tunnel.

the shortest path for a car by an OSM-based routing service, such as pgRouting or OpenRouteService. It is also possible to see that no vehicle takes this route and only pedestrians take the (indoor) route in order to get to the same ending point from the same starting point. This can also be another approach to infer or check path tags.

Another example of such an approach, that is analyzing user trajectories to recognize errors and bugs, is finding invalid connections (such as invalid motorway junctions). Invalid connections, junctions, or intersections can be detected if tracking data show no turning, short stop (while waiting for clear conditions to turn or due to a traffic lights at that point) or noticeable change in speed at or near such points (ignoring small deviations from average due to traffic delays). For example, based on travel behavior and patterns recognized over tracking data, it is possible to identify parking spaces, recognize wrongly tagged roads by finding the average speed of trajectories and also one-way “dead-end” roads (cul-de-sacs), which can be detected using headings of movements showing travelers moving in both directions. The underlying information extracted from the raw trajectories is, for example, based on type of travel mode and movement changes, from which higher level knowledge about patterns may be extracted.

In order to recognize such patterns and rules in data, an inference engine was developed as an ArcGIS add-in to store, visualize and analyze trajectories and then infer rules and patterns using spatio-temporal data mining techniques. Tracking data, which have been captured over a period of two month using a mobile app installed on volunteers’ mobile devices, was analyzed based on

techniques such as clustering, classification, and regression. In order to extract patterns and rules, the density, consistency, and frequency of the input trajectories are important. Depending on the area of coverage, the number of trajectories and the time interval in which trajectories are captured can vary significantly.

In order to analyze data with this approach, firstly anomalies and abnormalities are detected and excluded from input data-sets using statistical (spatial and temporal) analysis. Then, using clustering, classification methods and rule association approaches clusters of data are identified. At this stage, criteria such as spatio-temporal and topological relationships, speed, stops, and corresponding times are considered from which rules and patterns may be recognized.

Rules and patterns are used to detect anomalies and abnormalities, which can be considered as potential errors and bugs in OSM data. Potential errors and bugs are stored in a spatial data-set categorized into feature classes referring to specific groups of errors, such as invalid connections feature class, wrongly tagged feature class.

Finding errors and bugs using inferred rules and learnt patterns is more likely to have real-world results since the rules are based on actual movements while simultaneously the method is potentially a more dynamic approach to crowd sourcing of transport and navigation patterns.

The next section discusses OSM data quality and explores aspects of OSM data quality. Section three proposes a crowd source-based approach to detect potential errors of OSM data. In this approach, trajectories of

movement are analyzed to extract some patterns and rules, which help to detect anomalies and errors within OSM data. Finally in the fourth section, an implemented Arc GIS add-in which is served by the same data mining module is illustrated in support of the proposed approach.

2. OpenStreetMap data quality

2.1. OpenStreetMap data quality aspects and issues

The two-way interaction between users and providers of information in Web 2.0 has had a revolutionary effect on geospatial data exchange, including Web mapping (Haklay, Singleton, and Parker 2008). As a result, Volunteered Geographical Information (VGI) is a new source of information in which there is no definite traditional boundary between the authoritative map producers and the public map consumers (Goodchild 2007). Almost equally, VGI has also been referred to in the literature as crowd-sourced geographical data (Goodchild and Glennon 2010). OpenStreetMap (OSM) is one of the most prominent examples of crowd-sourced VGI. OSM, which was started in 2004 as a project (and in 2006 as a foundation), has attracted over two million registered users at the time of writing (OSM Wiki [n.d.](#)). OSM users may freely map any area of the world in a Web 2.0 collaborative manner, and the produced maps become instantly available for free public access all around the world. Users may map the world using a variety of technique, such as using GPS traces or their local knowledge assisted with some aerial imagery (Haklay and Weber 2008). Moreover, the unrestricted usage of key-value pairs for tagging features provides an excellent means of customized annotations suitable for thematic applications. A complete review of OSM recent developments is available (Neis and Zielstra 2014).

Although OSM is rapidly growing in content and contributors, its credibility has been one of the main concerns for authoritative users. The belief that it is made by amateurs is perceived to limit trust in the value of this free data source within the traditional GIS community. OSM data quality has been the main concern regarding its reliability among map consumers, especially for these authoritative consumers (Flanagin and Metzger 2008; Fonte et al. 2015). The quality of OSM data (for different geographical features and/or contexts) is extensively studied and analyzed in the literature. A detailed review is not in the scope of this paper. More detailed studies are available, (Barron, Neis, and Zipf 2014; Helbich et al. 2012; Mondzech and Sester 2011; Ludwig, Voss, and Krause-Traudes 2011; Kounadi 2009; Fan et al. 2014; Mooney, Corcoran, and Winstanley 2010; Salk et al. 2015; Herfort, Eckle, and Zipf 2015; Ali et al. 2014; Hashemi and Abbaspour 2015).

There are, however, some OSM-specific quality characteristics (as reviewed by Mooney, Corcoran, and Winstanley 2010) that come to our focus. In part, this relates to the free and open nature of feature attribution in OSM. Taxonomically, each feature in OSM can have an unlimited number of attributes in a key-value pair free-text format. The OSM community has documented a list of key-value pairs that can be used to describe a real-world feature (OSM Wiki [n.d.](#)). These can be used as a reference for quality metrics; however, the list may change over time and users are free to use their own tags to elaborate particular feature attributes to be used in customized rendering tools.

OSM data quality can be checked using two main approaches: comparing OSM data with authoritative resources, and rule-based (manual or automatic) self-detection. In the latter, the rules may be either user specified or extracted from the data. The next subsection explains each approach with implemented examples.

2.2. OpenStreetMaps data validation techniques and methods

As described above, there are three main categories of OSM quality control: (a) comparing OSM data against authoritative spatial data, (b) user and rule-based checking, and (c) crowd-source rule and pattern extraction for rule-based checking. This sub-section explains the first two categories in more detail and the next section explains the last one providing an application and implementation example.

2.2.1. Comparing OSM data with “authoritative” resources

The success, openness, and free availability of OSM has made it a very good examination ground for researchers to study the different aspects of collaborative mapping characteristics, such as comparative accuracy and completeness analysis (Haklay 2010) and patterns of data collection (Haklay et al. 2010) (Mooney and Corcoran 2011).

Existing research on OSM quality statistically compares an OSM snapshot with reference maps in order to assess OSM's overall accuracy and/or completeness. For example, the positional accuracy of the OSM features is evaluated by matching them with the UK national Ordnance Survey mapping (Ather 2009; Haklay 2010; Kounadi 2009). Similar investigations have been undertaken by Zielstra and Zipf (2010), Neis, Zielstra, and Zipf (2013), and Barron, Neis, and Zipf (2014). Regarding attribution accuracy and/or completeness, the naming of the UK road network has been compared to that by the Ordnance Survey (Pourabdollah 2014).

2.2.2. User and rule-based checking

Internal data-validation methods have been developed, such as node spacing on specified feature polygons in an OSM snapshot (Mooney and Corcoran 2011) and other online and desktop tools (Haklay 2010; Haklay et al. 2010; JOSM n.d.; Kounadi 2009). These methods and tools can either be based on user-checking or on predefined automated rules which can detect and correct bugs and errors. For example, Pourabdollah et al. (2013) introduced 17 rules to a system that manually checks OSM's geometry and attribute data quality. The system can find some specific types of bugs and errors such as wrongly tagged bridges/tunnels, invalid motorway connections and one-way dead-end roads. Such rules are mainly based on logical assumptions and mapping agencies standards. For example, two roads crossing at the same level must have an intersection. However, another controlling process is typically also needed to check the real-world validity of identified errors and bugs. If rules were based on users' travel behavior, rather than manual or even automated rule-based checking, detected errors and bugs would be more likely to be correctly identified and less labor-intensive checking would be needed.

There are some enhanced error discovery and/or error reporting tools available for OSM, as well as some other tools to validate or fix the errors locally before feeding it into OSM. There are also a number of community-supported error-reporting tools, in which OpenStreetBugs (OpenStreetBugs n.d.) and MapDust (Scobbler GmbH 2012) are two known examples.

In addition, online or desktop tools exist ranging from pre- to post-validation tools. Error detecting tools such as KeepRight (2012), OSMOS (n.d.), and OSM Inspector (2012) perform the automatic analysis of OSM data and visualize the detected errors on slippy maps. KeepRight nicely visualizes the detected bugs but the front page visualization is the only means of access to the detected errors. OSMOS does the similar job in France and OSMInspector does the same in Germany. JOSM Validator is an integrated part of JOSM (n.d.), the Java-based OSM editor. The program runs a series of pre-entry validation tests on the data before they can be uploaded to OSM. A full list of OSM quality assurance tools is available (OSM Wiki n.d.).

The next section describes crowd-source rule and pattern extraction for OSM quality assurance purposes. In particular, it focuses on finding errors in OSM data based on spatial knowledge extracted from anonymous tracking data of users.

3. Validation using crowd-sourced trajectory mining

There is another approach which benefits from the advantages of crowd-source data capture, although it is not based on the direct contribution of users. Users

do not edit OSM data directly, but it is edited based on knowledge, patterns, and rules extracted from movements. Contributors provide to a centralized repository their movement trajectories and an inference system finds patterns in the data and derives inherent rules which can be used to assess and improve the quality of OSM data. Such patterns and rules may highlight anomalies and unusual data which may be errors and which can then be manually examined or in some cases automatically corrected.

For example, if a cluster of trajectories, classified as "driver" based on average speed and map-matched features, shows travel in both directions, while the OSM data attribute for the matching feature is a one-way street, then it is possible to identify a potential error. However, this error and bug detection approach uses a conservative strategy; the robustness of the rules and the reliability of the results highly depend on the application requirements, the spatial, temporal coverage, and size of training and test data-sets. Any bug and error identified by this approach is highlighted for more investigation.

Another example of such an approach is finding invalid connections (such as invalid motorway junctions). Invalid connections, junctions, or intersections can be detected if tracking data show no turning, short stop (while waiting for clear conditions to turn or due to a traffic light at that point) or noticeable change in speed at or near such points. There are many more possibilities for inference-based error and bug detection rules that could be specified from analysis of trajectory data to check and improve OSM data quality.

Rules and patterns can be applied to detect anomalies and abnormalities, which potentially indicate errors and bugs. They can be stored in a spatial data-set, categorized into feature classes referring to specific groups of errors, such as invalid connections, wrongly tagged feature class, newly constructed features which have not been added to OSM.

Finding errors and bugs using inferred rules and learnt patterns from crowd-source data is more likely to have valid real-world results since the rules are based on actual behavior. This offers a potentially more dynamic approach to crowd sourcing of transport and navigation patterns. In addition, it extracts rules and patterns from trajectories so is more adaptable to the domain, since many pieces of information can be potentially extracted from trajectories. In contrast, a user-quality-control approach, where contributors directly edit OSM data, users edit based on their own knowledge, and understanding which might not be correct or up-to-date. This can cause many contradictory edits from different users, who have different understandings of meaning of each tag and attribute, different knowledge of spatial accuracy of a feature, and so on. By using automatically captured trajectories, it is possible to extract such knowledge and take action (edit, delete, or insert a feature)

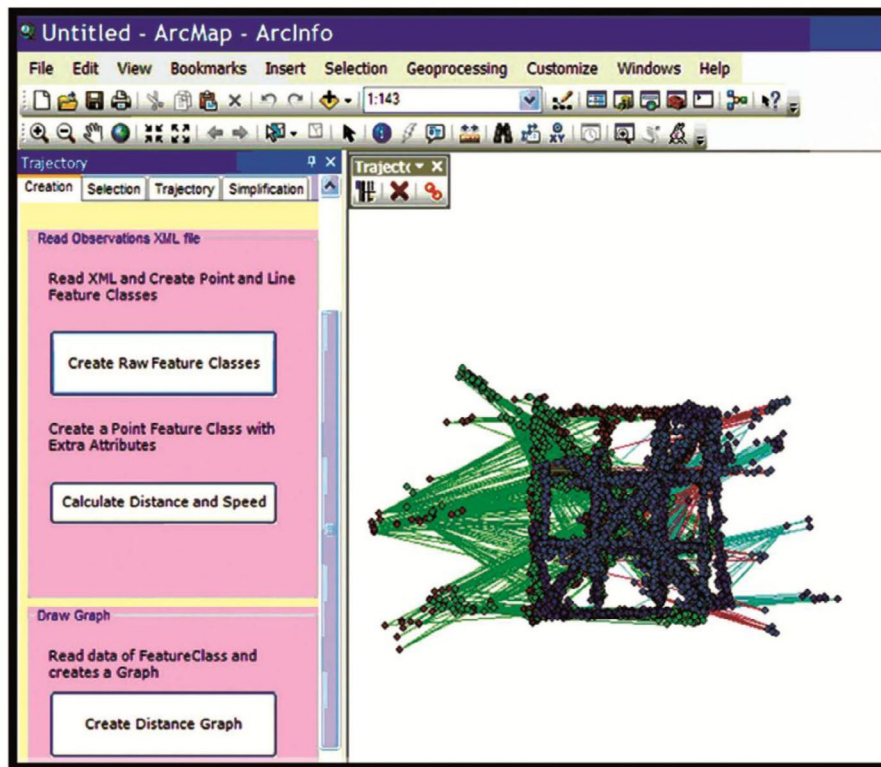


Figure 2. Trajectory analyzer dockable window (developed ArcGIS add-in).

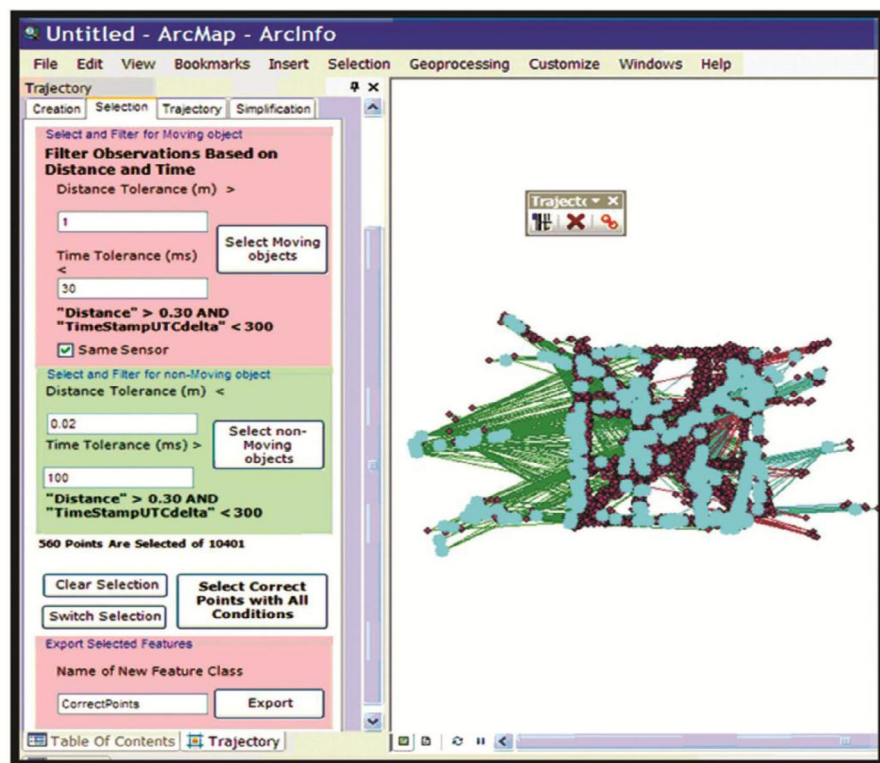


Figure 3. Noise/error filtering and stay point detection tab.

where a large enough sample of trajectories support the inferred knowledge and pattern.

In order to extract rules and patterns, we use several spatio-temporal data mining techniques. There are some preconditions that should be satisfied to make sure that the results are valid and are not based on incomplete input data sets. If the input data-set is small (for example, in terms of the number of samples and the extent and density/frequency of trajectories with respect to space and time), then outputs (the rules and patterns) may be valid only for a specific situation, area, and time interval. The level of robustness of output knowledge is strongly correlated with input sample size. In this regard, for successful pattern recognition and rule extraction, there are some rules-of-thumb:

- (i) The sample size of trajectory data has to be large, otherwise the training and control/test process may not find all patterns contained in the data.
- (ii) The sample data should be dense enough to give complete spatial coverage.
- (iii) The sample data should be frequent enough to give complete temporal coverage (different days, times, weekends, seasons, and so on)
- (iv) The sample data should cover all (at least most) travel behaviors and modes. This helps spatio-temporal data mining to exclude anomalies and exceptions, and also it helps to find clusters and classes which share common patterns easier and with a greater level of certainty.

With adequate input data, it is possible to apply spatio-temporal data mining techniques to identify the patterns, clusters, and rules, which can model the traveler behaviors and movements.

The first step is capturing the trajectory. This can use many different positioning and tracking technologies and methods including Global Navigation Satellite Systems (GNSS), Wireless Local Area Network (WLAN), Radio Frequency Identification (RFID), cameras, mobile networks, Inertial Navigation Systems (INS), Bluetooth networks, tactile floors, and UltraWide Band (UWB) (Basiri et al. 2015).

We implement this process using an inference engine, developed as an ArcGIS add-in. Tracking data, which have been captured over a period of two months using a mobile app installed on mobile devices, was analyzed based on the algorithm.

In order to show that there is a strong correlation between the reliability of the output from data mining and the bounding box (both spatially and temporally) in which the trajectory data are located, two sets of trajectory data for different cities (Hanover in Germany and Maynooth in Ireland) were captured. Trajectories were captured over two months (July 2013 to August 2013) using a mobile app which can be downloaded from servers at the Institute of Cartography and Geoinformatics (IKG) at the Leibniz University of Hanover and the

National University of Ireland, Maynooth (NUIM). Positional data were captured using GPS and mobile networks (usually when user is moving outdoors) or it can be captured from QR codes affixed to most of the major turning points and important features (specially for indoor localization) using a QR code reader app (Basiri, Amirian, and Winstanley 2014; Basiri et al. 2016). It is also possible to capture the trajectory of movement from a network of ceiling mounted cameras (such as CCTVs).

There are two pre-processing stages on the data: anonymity control and noise filtering/error exclusion.

Due to privacy and data protection issues, it is highly important to anonymize the data especially when the data are from CCTV cameras (Gidofalvi, Huang, and Pedersen 2007). Therefore, tracking data are stored without any reference to a user's identification. There are various anonymisers (Chow and Mokbel 2011). This paper uses the K-anonymity program (Kalnis et al. 2007), which is trusted third party software often used on tracking data (Chow and Mokbel 2011). The data are stored in centralized systems or on decentralized peer devices (Ghinita, Kalnis, and Skiadopoulos 2007). The anonymizer removes the ID of the user and cloaks the exact user location in the spatial database.

Anonymized trajectories can have some points that are not perfectly accurate or in some cases even valid. Such errors and noises should be filtered in advance to minimize the invalid results at the end of the data mining process. It is very important to remember this phase is being carried out to filter the errors and noises and it is not to exclude the abnormalities and anomalies, as they might be helpful for some applications and scenarios. The noises and errors are in the data due some reasons including poor and multipath positioning signals or tracking the reflection (for example on windows) rather than actual location of the users by CCTV camera. In order to detect and exclude noises and errors, there are some methods available, such as Kalman and Particle filtering and mean (or median) filters, described and reviewed by Lee and Krumm (2011).

This application, however, used a heuristic-based method, as the other filters replace the noise/error in the trajectory with an estimated value and this may have a significant impact on the output of trajectory mining (i.e. the recognized patterns and rules). It calculates the distance and the travel time between each consecutive couple of points in the trajectory and then the travel speed for each segment can be easily calculated. Then it is possible to find out the segments whose travel speeds are larger than a threshold (for example, 360 km/h). If the travel mode for each trajectory is also identifiable (identified by the contributors or based on statistical methods which can find of the consecutive segments with the almost the same average speed classifiable into three classes of pedestrian, car/bus/ train, and bicycle),

then it is possible to have different thresholds depend on the travel mode (a pedestrian cannot walk faster than 20 km/h.)

This paper only uses the travel modes specified by the contributors and simply ignore the possibility of error/noise detection using the statistical methods. This is due to some transitional segments (e.g. from pedestrian to car and then again to pedestrian mode) or some anomalies (which are not due to errors or noises) that might be removed if their spatial characteristics and relationships with surrounding spatial features (i.e. map matching) is not considered. Such segments need to be carefully kept for the next steps of trajectory data mining process as they potentially can have valuable information.

As it is previously mentioned, an ArcGIS add-in has been developed, as shown in Figure 2, to visualize, process and analyze the input trajectory data. As illustrated in Figure 2, a trajectory analyzer in a dockable window is available to ArcMap. The first tab creates a feature class by reading the input XML file of the recorded points (i.e. GNSS logs, scanned QR-Codes, and mobile cell-ID locations). It can add two columns to the created feature class which calculate distance between each adjacent point pair (the length of each segment) and the speed of movement of the user passing that segment. The travel speed is compared to the threshold (depending on the travel mode and if the travel mode is not specified by the contributors, it is being set to 360 km/h, however, this value is a user-defined value and can be easily changed, as it shown in Figure 3). The travel speed is being stored as an attribute data to each segment as later on it is being used for rule association and also pattern recognition steps.

In addition to the classification of segments, this function also helps to exclude redundant data. For example, it is possible to find points where the user has been stationary (speed close to zero) and replace them with a single point with a description of the time interval during which the user's speed was zero.

By calculating the correlation between trajectory data, it is possible to discover other modes of classification. Spatio-temporal clustering helps to identify such classes and to discover underlying rules and patterns through identifying parameters which are highly correlated and extracting rules. Evolution rules, explained in Section 3, are applied at this stage through functions on the selection tab to discover spatial and temporal rules. Figure 2 shows clusters and classes of vertices of trajectories depending on area of search, buffer threshold which limits numbers of trajectories with the same types (modes) within an area. There is a buffering since there are always inaccuracies due to positioning technologies. In addition to buffer area, there is buffering for temporal aspect of input data which limit time interval when trajectories can be matched. Because of the large number of input features, spatial and temporal thresholds are used to cluster and identify associations. Such thresholds

depend on the density, frequency, and the nature of the input data. They can be changed using this tab according to experts' comments and recommendations.

Having the data anonymized and error/noise excluded, then trajectories are ready to go through the next stage and be pre-processed. Pre-processing stage is to make them easier to store (using segmentation, compression, and simplification techniques) and also semantically more understandable (by identifying the stay points and using the map matching techniques). The pre-process step is actually based on the fact that all the points in a trajectory are not equally important and meaningful (Zheng 2015). The pre-process step, including stay point detection, trajectory segmentation, trajectory simplification, and compression make the trajectories ready to be stored and retrieved in a more efficient way. In addition to the efficiency, trajectory segmentation and stay point detection make the trajectories and some of the points semantically meaningful and easier to interpret. Such stay points can be used in recommended places to visit (Basiri, Amirian, and Winstanley 2014; Luo et al. 2013), estimate the actual travel time and petrol consumption (Shang et al. 2014).

Stay points refer to the locations where users/contributors have stayed for a while. The stay points can be simply identified if the location of the user is not changing over a period of time. However, due to positioning services' inaccuracy and errors (Pang et al. 2013), it is commonly happens that the user stays stationary for a while but the positioning technology generate different readings (Zheng 2015). In order to detect such stay points/area, there are several algorithms and methods; Basiri et al. (2014) proposed an approach that checks the travel speed for each segment and if it is smaller than a threshold, then the average of these two points are replaced/stored as the "stay point". It is possible to put distance and temporal interval threshold separately, instead of speed of each segment (i.e. if the distance between a point and its successor is larger than a threshold and also the time span is larger than a given value) as Li et al. (2008) proposed. Yuan et al. (2015) proposed using the density-clustering algorithm to identify the stay points.

This paper uses this approach in order to identify the speed clustering algorithms to identify the stay points. This can be viewed as the combination of segments' speed threshold and the density-clustering algorithms.

Abnormalities and anomalies are detected (using statistical analysis), then some clusters are identified (using association rules between average speed between two points, number of stops, duration of stops, spatio-temporal topological relationships between trajectories and available features on the maps). Then rules and patterns can be recognized using relevant parameters and criteria, including speed, spatial, and temporal correlations between segments and trajectories and

also some trajectory matching algorithms. Some of the examples of such rules and patterns are listed in below:

If the average speed of movement is more than 50 km/h and the trajectory is matched by the street network, then the travel mode is car; if instead the trajectory is matched with bus lines and there are stops (the speed for a short period of time, becomes zero) at the bus stops, then the travel mode may be bus. Such rules can be used in the phase of “recognizing user’s current situation”.

The inference can be more complicated than simple if-then rules. It is possible to find the pattern over input data-set using data mining techniques. By analyzing input trajectories captured from different types of users, it is possible to automatically find some similarities and patterns.

There are some research projects and applications focusing on finding patterns of movements over users’ trajectories, such as detecting travel modes (Feuerhake 2012; Zhang, Dalyot, and Sester 2013), group movement patterns (Dodge, Weibel, and Lautenschütz 2008), unusual behavior detection from trajectory analysis (Kuntzsch and Bohn 2013), etc. However, they focus on finding patterns to find out more about movement behavior and interpretation of them. In addition, some reference spatial data, such as maps, are used as another set of input data (or available rules) in the process of pattern recognition over trajectory data. This shows that the accuracy and reliability of reference maps are assumed unquestionable, however, this is not always true especially for crowd-source data since many contributors are not aware of impact of spatial data quality aspects. This paper focuses on identifying patterns and rules over movements’ trajectories in order to identify OSM data bugs and errors.

Data mining techniques, which have been implemented in some pattern recognition projects, are based on static approaches (conventional data mining techniques), which are not fully compatible with the spatial and temporal aspect of trajectory data. Also in some research projects (Monreale et al. 2009; Yavas et al. 2005) where dynamic data mining has been applied, spatial and temporal aspects of input data were considered separately. This approach does not include spatio-temporal relationships which can help to identify some other rules.

The problems with most spatial and temporal (not spatio-temporal) data mining techniques, which have been used for pattern recognition, are given as the following:

- Spatio-temporal topological relationships are ignored. The spatial relations, both metric (such as distance) and non-metric (such as topology, direction, shape, etc.) and the temporal relations (such as before and after) are information bearing and therefore need to be considered in the data mining methods. Also, some spatial and temporal

relations are implicitly defined and they are not explicitly encoded in a database. These relations should be extracted from the data. However, there is always a trade-off between pre-computing them before the actual mining process starts (eager approach) and computing them on the fly when they are actually needed (lazy approach). Moreover, despite much formalization of space and time relations available in spatio-temporal reasoning, the extraction of spatial/temporal relations implicitly defined in the data introduces some degree of certainty that may have a large impact on the results of the data mining process.

- Working at the level of stored data, that is, geometric representations (points, lines, and regions) for spatial data or time stamps for temporal data, is often undesirable. Therefore, complex transformations are required to describe the units of analysis at higher conceptual levels, where human-interpretable properties and relations are expressed.
- Spatial resolution or temporal granularity can have a direct impact on the strength of patterns that can be discovered in the data-sets. General patterns are more likely to be discovered at the lowest resolution/granularity level. On the other hand, large support is more likely to exist at higher levels of resolution. To have a better support and also have more clusters, higher level of resolution and granularity are needed. In this stage, it worth to mention that, lack of spatial and temporal accuracy and precision of input data can be compensate, up to some extent, by number of input trajectories. So for large enough data-sets, less accurate trajectories can let us have appropriate results as if accurate data were available.

In order to consider spatio-temporal relationships in the data mining process, a spatio-temporal database, which can store all required aspect of spatio-temporal objects, should be firstly generated. This helps to use, modify, and analyze different characteristics and relationships between/of trajectory data. Then using such a database and having large enough input datasets, it would be possible to find similarities, clusters, classes, anomalies, etc. and finally find rules and patterns contained within the movement trajectory clusters.

The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data-set. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data-set. This is called overfitting. To overcome this, the evaluation uses a test set of data on which the data mining algorithm was not trained. The learned patterns are applied to this

test set and the resulting output is compared to the desired output. The accuracy of the patterns can then be measured from how many are correctly classified. A number of statistical methods may be used to evaluate the algorithm, such as ROC curves.

If the learned patterns do not meet the desired standards, then it is necessary to re-evaluate and change the pre-processing and data mining steps. If the learned patterns do meet the desired standards, then the final step is to interpret the learned patterns and turn them into knowledge.

However, data mining techniques are very well developed, in order to have a better pattern recognition process and consider all aspects of trajectory data, it is highly recommended to apply spatio-temporal data mining techniques not to have problem described in the last subsection in last subsection. In this case, spatio-temporal relationships are also considered in this process.

Spatio-temporal data mining includes a set of computational techniques for the analysis of large spatio-temporal databases. Both the temporal and spatial dimensions add substantial complexity to data mining tasks.

First of all, the spatial relations, both metric (such as distance) and non-metric (such as topology, direction, shape, etc.) and the temporal relations (such as before and after) are information bearing and therefore need to be considered in the data mining methods (Basiri and Malek 2014).

Second, some spatial and temporal relations are implicitly defined and they are not explicitly encoded in a database. These relations should be extracted from the data. However, there is always a trade-off between pre-computing them before the actual mining process starts (eager approach) and computing them on-the-fly when they are actually needed (lazy approach). Moreover, despite much formalization of space and time relations available in spatio-temporal reasoning, the extraction of spatial/temporal relations implicitly defined in the data introduces some degree of certitude that may have a large impact on the results of the data mining process.

Third, working at the level of stored data, that is, geometric representations (points, lines, and regions) for spatial data or time stamps for temporal data, is often undesirable. Therefore, complex transformations are required to describe the units of analysis at higher conceptual levels, where human-interpretable properties and relations are expressed. Fourthly, spatial resolution or temporal granularity can have direct impact on the strength of patterns that can be discovered in the data-sets.

As discussed by Abraham and Roddick (1998), the forms that spatio-temporal rules may take are extensions of their static counterparts and at the same time are uniquely different from them. Five main types can be identified:

- **Spatio-Temporal Associations.** These are similar in concept to their static counterparts as described by Agrawal, Imielinski, and Swami (1993). Association Rules are of the form $X \rightarrow Y$ ($c\%$, $s\%$), where the occurrence of X is accompanied by the occurrence of Y in $c\%$ of cases (while X and Y occur together in a transaction in $s\%$ of cases).
- **Spatio-Temporal Generalization.** This is a process whereby concept hierarchies are used to aggregate data, thus allowing stronger rules to be located at the expense of specificity. Two types are discussed in the literature; spatial-data-dominant generalization proceeds by first ascending spatial hierarchies and then generalizing attributes data by region, while nonspatial-data-dominant generalization proceeds by first ascending the spatial attribute hierarchies. For each of these different rules may result.
- **Spatio-Temporal Clustering.** While the complexity is far higher than its static, non-spatial counterpart the ideas behind spatio-temporal clustering are similar – that is, either characteristic features of objects in a spatio-temporal region or the spatio-temporal characteristics of a set of objects are sought (Ng 1996; Ng and Han 1994).
- **Evolution Rules.** This form of rule has an explicit temporal and spatial context and describes the manner in which spatial entities change over time. Due to the exponential number of rules that can be generated, it requires the explicit adoption of sets of predicates that are usable and understandable. Example predicates include Follows, Coincides, Parallels, and Mutates (Allen 1983; Freksa 1992; Hornsby and Egenhofer 1998).
- **Meta-Rules.** These are created when rule sets rather than data-sets are inspected for trends and coincidental behavior. They describe observations discovered among sets of rules. For example, the support for suggestion X is increasing. This form of rule is particularly useful for temporal and spatio-temporal knowledge discovery.

In order to extract patterns and rules of movements, all input trajectory data are randomly divided into two feature classes. The first one, which is called, training data, is used for pattern recognition and rule learning purposes. Another set of input data, which is called control data, is used to control how the learnt rules and recognized patterns fit into this set of data. After analyzing and finding patterns on the training data, the inference system will apply the extracted patterns on the control data to see how similar input control data and estimated results are. If very similar, it is possible to infer that a pattern was discovered and any new data can be analyzed using that pattern.

The clusters and classes which can be generated using above-mentioned techniques are used to generate rule sets. The final step of knowledge discovery from data is to verify that the patterns produced by the data mining algorithms occur in the wider data-set. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data-set.

If the learned patterns do not meet the desired standards, then it is necessary to re-evaluate and change the pre-processing and data mining steps. If the learned patterns do meet the desired standards, then the final step is to interpret the learned patterns and turn them into knowledge.

In this paper, since prior knowledge about the input data was included (such as a reference map or additional spatial data), it was decided to evaluate the correctness and logic of output patterns and rules using standards, “common sense” rules-of-thumb and expert comments as well as control data test. However, there is a need to compare the results of this approach with other approaches results to evaluate them. One of the early inferred rules and patterns is about identifying the travel mode using speed and behavior of movement. Based on speed of movement, it is possible to classify data into the four categories of pedestrian, bicycle, wheelchair, and vehicle. Using patterns of movement, it is possible to find some rules which distinguish between public transportation and cars—public transportation stops regularly at very specific points with very low correlation to time, that is whenever a vehicle arrives at station, it usually stops. Such rules and patterns should be confirmed by control data. However, if no reference data is available, expert comments, logical rules, and standard specifications are also part of this process.

Using speed and pattern of movement it is possible to identify junctions, bus lanes, pedestrian-only routes, one-way roads, no-U turns, and so on. Figure 3 shows results of clustering and classification methods from raw trajectory data. Comparing such clusters and classes with OSM data makes it possible to identify incompatibilities and incompleteness in the OSM data. However, there are two issues with this process: the accuracy of the inferred features, such as junctions and stops, and the reliability of the inferred rules. Since input data, that is the trajectories, are captured from positional data, they always suffer from inaccuracy and uncertainty. Although there are many trajectories used to infer a single feature and so the impact of inaccuracies is lower, there can still be a problem to identify the matching feature in OSM since they will never be spatially identical. Regarding the reliability of inferred rules, there is also another control process before changing the OSM data. Another feature class is created to store mismatched features for further investigation.

4. Conclusions and future work

OpenStreetMap (OSM) data are widely used, however, its reliability is still under question since many contributors are not geospatial data experts. There are some systems and applications dedicated to finding bugs and errors of OSM data. Such systems can find errors using user-checking or predefined rules but they all need an extra controlling process to check the real-world validity of any apparent errors and bugs. If rules were based on the users’ travel behavior, rather than standard specifications, errors and bugs would more likely be identified correctly. This paper focuses on finding bugs and errors using patterns and rules extracted from tracking data. In order to recognize such patterns and rules, an inference engine was developed as an ArcGIS add-in which can store, visualize, and analyze trajectories and then infer rules and patterns using spatial data-mining techniques. Using such rules, potential bugs and errors in OSM data can be identified and stored for further investigation.

Funding

This research was supported financially by EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-technology Positioning Professionals) [grant number 316528].

Notes on contributors

Anahid Basiri is a Marie Curie Experienced Research Fellow at the University of Nottingham. Ana works on future applications and markets of Location-Based Services (LBS). She studies current LBS applications challenges including privacy and accuracy and identifies potential solutions and trends to forecast future markets of LBS. She got her PhD in 2012 in the field of Geospatial Information Systems (GIS). Her research interest includes LBS applications and specifically navigation services, GIS, spatial uncertainty, data analysis, application development.

Mike Jackson is an Emeritus Professor at the Nottingham Geospatial Institute, University of Nottingham. His research interests are spatial data infrastructures, crowd-sourcing and geo-intelligence. He is a non-executive director of the Open Geospatial Consortium and past President of the Association of Geographic Laboratories in Europe (AGILE).

Pouria Amirian is a Principal Scientist in Data Science and Big Data at the Ordnance Survey, UK. He is also a research associate with the University of Oxford. He has been involved in several big data and data science projects. He is interested in management and distributed analysis of geospatial big data.

Amir Pourabdollah received his PhD in Computer Science from The University of Nottingham, UK in 2009, with mixed backgrounds in telecommunications, software engineering and GIS. Since then, he is a research fellow in Nottingham Geospatial Institute and the School of Computer Science at the same university. His recent research projects have been on topics such as geospatial data modeling, linked-data, open tools/standards/data, geospatial workflows, Web mapping, crowd-sourced GIS, uncertainty management using artificial intelligence and fuzzy logic techniques.

Monika Sester is a Professor and the Head of the Institute of Cartography and Geoinformatics at Leibniz Universität Hannover. Her research interests lie in data interpretation, analysis and visualization. A current focus is on VGI data and the challenges and benefits they bring about for data management, analysis and interpretation.

Adam Winstanley is the Head of the Department of Computer Science, Maynooth University. He was Co-Principal Investigator on the StratAG project into Advanced Geomatics 2007–2013 funded by Science Foundation Ireland and has research interests in navigation, location-based systems, and data science applied to spatial and historical data.

Terry Moore is the leader of the Nottingham Geospatial Institute, *University of Nottingham* and has research interests in the development of positioning and navigation technologies with applications across many sectors of aerospace and transportation. He is currently the Vice President of the Royal Institute of Navigation.

Lijuan Zhang is a postdoctoral fellow at the Technion-Israel Institute of Technology. She received her PhD in Geoinformation from Leibniz University Hannover in 2014. Her research interests lie in the fields of crowd-sourced volunteered geographic information (VGI), spatio-temporal data mining and location based services.

ORCID

Anahid Basiri  <http://orcid.org/0000-0002-2399-1797>

Adam Winstanley  <http://orcid.org/0000-0003-0817-0832>

References

- Abraham, T., and J. F. Roddick. 1998. "Opportunities for Knowledge Discovery in Spatio-Temporal Information Systems." *Australasian Journal of Information Systems* 5 (2): 3–12.
- Agrawal, R., T. Imielinski, and A. N. Swami. 1993. "A Mining Association Rules between Sets of Items in Large Databases." In *Proc. ACM SIGMOD Conf. on Management of Data*, 207–216. New York, NY: ACM.
- Ali, A. L., F. Schmid, R. Al-Salman, and T. Kauppinen. 2014. "Ambiguity and Plausibility: Managing Classification Quality in Volunteered Geographic Information." *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 143–152. New York, NY: ACM.
- Allen, J. F. 1983. "Maintaining Knowledge about Temporal Intervals." *Communications of the ACM* 26 (11): 832–843.
- Amirian, P., A. Basiri, G. Gales, A. Winstanley, and J. McDonald. 2015. "The Next Generation of Navigational Services Using OpenStreetMap Data: The Integration of Augmented Reality and Graph Databases." In *OpenStreetMap in GIScience – Experiences, Research, and Applications*, edited by J. Jokar Arsanjani, A. Zipf, P. Mooney, M. Helbich, 211–228. Berlin Heidelberg: Springer.
- Arsanjani, J. J., P. Mooney, A. Zipf, and A. Schauss. 2015. "Quality Assessment of the Contributed Land Use Information from OpenStreetMap versus Authoritative Datasets." In *OpenStreetMap in GIScience, Experiences, Research, and Applications (LNCS)*, edited by J. Jokar Arsanjani, A. Zipf, P. Mooney, M. Helbich, 37–58. Berlin Heidelberg: Springer.
- Ather, A. 2009. "A Quality Analysis of OpenStreetMap Data." Master's Thesis, University College London.
- Barron, C., P. Neis, and A. Zipf. 2014. "A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis." *Transactions in GIS* 18 (6): 877–895.
- Basiri, A., and M. R. Malek. 2014. "Providing Relevant Information in an Ambient Services Using Service Requester's "Logical Area"." *Journal of Ambient Intelligence and Humanized Computing* 5 (4): 539–549.
- Basiri, A., P. Amirian, and A. Winstanley. 2014. "The Use of Quick Response (QR) Codes in Landmark-Based Pedestrian Navigation." *International Journal of Navigation and Observation* 2014: 897103, 7 pages.
- Basiri, A., P. Petolta, P. F. Silva, E. S. Lohan, T. Moore, and C. Hill. 2015. "Indoor Positioning Technology Assessment Using Analytic Hierarchy Process for Pedestrian Navigation Services." In *International Conference on Localization GNSS*. Gothenburg, Sweden: IEEE.
- Basiri, A., P. Amirian, A. Winstanley, S. Marsh, T. Moore, and G. Gales. 2016. "Seamless Pedestrian Positioning and Navigation Using Landmarks." *Journal of Navigation* 69 (01): 24–40.
- Chow, C. Y., and M. F. Mokbel. 2011. "Privacy of Spatial Trajectories." In *Computing with Spatial Trajectories*, edited by Y. Zheng, X. Zhou, 109–141. Berlin Heidelberg: Springer.
- Dodge, S., R. Weibel, and A.-K. Lautenschütz. 2008. "Towards a Taxonomy of Movement Patterns." *Information Visualization* 7: 240–252.
- Fan, H., A. Zipf, Q. Fu, and P. Neis. 2014. "Quality Assessment for Building Footprints Data on OpenStreetMap." *International Journal of Geographical Information Science* 28 (4): 700–719.
- Feuerhake, U. 2012. "Prediction of Individual's Movement Based on Interesting Places." *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1–2: 31–36.
- Flanagin, A., and M. Metzger. 2008. "The Credibility of Volunteered Geographic Information." *GeoJournal* 72 (3–4): 137–148.
- Fonte, C. C., L. Bastin, L. See, G. Foody, and F. Lupia. 2015. "Usability of VGI for Validation of Land Cover Maps." *International Journal of Geographical Information Science* 29: 1269–1291.
- Freksa, C. 1992. "Using Orientation Information for Qualitative Spatial Reasoning." *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, LNCS* 639: 162–178.
- Ghinita, G., P. Kalnis, and S. Skiadopoulos. 2007. "MobiHide: A Mobile Peer-to-Peer System for Anonymous Location-Based Queries." *Advances in Spatial and Temporal Databases, Lecture Notes in Computer Science* 4605: 221–238.
- Gidofalvi, G., X. Huang, and T. B. Pedersen. 2007. "Privacy Preserving Data Mining on Moving Object Trajectories." In *Proceedings of the 8th IEEE International Conference on Mobile Data Management*, 60–68. Mannheim, Germany, May 7–11.
- Goodchild, M. F. 2007. "Citizens as Sensors: The World of Volunteered Geography." *GeoJournal* 69 (4): 211–221.
- Goodchild, M. F., and J. A. Glennon. 2010. "Crowdsourcing Geographic Information for Disaster Response: A Research Frontier." *International Journal of Digital Earth* 3 (3): 231–241.
- Haklay, M. 2010. "How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and

- Ordnance Survey Datasets." *Environment and Planning B, Planning and Design* 37 (4): 682–703.
- Haklay, M., and P. Weber. 2008. "OpenStreetMap: User-Generated Street Maps." *IEEE Pervasive Computing* 7 (4): 12–18.
- Haklay, M., A. Singleton, and C. Parker. 2008. "Web Mapping 2.0: The Neogeography of the GeoWeb." *Geography Compass* 2 (6): 2011–2039.
- Haklay, M., S. Basiouka, V. Antoniou, and A. Ather. 2010. "How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information." *The Cartographic Journal* 47 (4): 315–322.
- Hashemi, P., and R. A. Abbaspour. 2015. "Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept." In *OpenStreetMap in GIScience, Experiences, Research, and Applications (LNCS)*, edited by J. Jokar Arsanjani, A. Zipf, P. Mooney, and M. Helbich, 19–36. Berlin Heidelberg: Springer.
- Helbich, M., C. Amelunxen, P. Neis, and A. Zipf. 2012. "Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata." In *Proceedings of GI Forum 2012*, 24–33. July 5–8, Salzburg, Austria.
- Herfort, B., M. Eckle, and A. Zipf. 2015. "Towards Assessing the Quality of Volunteered Geographic Information from OpenStreetMap for Identifying Critical Infrastructures." In *Proceedings of the ISCRAM 2015 Conference*. May 24–27, Kristiansand, Norway.
- Hornsby, K., and M. Egenhofer. 1998. "Identity-Based Change Operations for Composite Objects." In *Eighth International Symposium on Spatial Data Handling*, edited by T. Poiker and N. Chrisman, 202–213. July, Vancouver, Canada.
- JOSM. n. d. "Java OSM Editor." Accessed April 12, 2014. <http://josm.openstreetmap.de>
- Kalnis, P., G. Ghinita, K. Mouratidis, and D. Papadias. 2007. "Preventing Location-Based Identity Inference in Anonymous Spatial Queries." *IEEE Transactions on Knowledge and Data Engineering* 19 (12): 1719–1733.
- KeepRight. 2012. "Data Consistency Checks for OSM." Accessed April 12, 2014. <http://keepright.at>
- Koukoletos, T., M. Haklay, and C. Ellul. 2012. "Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data." *Transactions in GIS* 16 (4): 477–498.
- Kounadi, O. 2009. *Assessing the Quality of OpenStreetMap Data*. Msc Geographical Information Science: University College of London Department of Civil, Environmental and Geomatic Engineering.
- Kuntzsch, C., and A. Bohn. 2013. "A Framework for on-Line Detection of Custom Group Movement Patterns." In *Progress in Location-Based Services Lecture Notes in Geoinformation and Cartography*, edited by J. M. Krisp, 91–107. Heidelberg: Springer.
- Lee, W. C., and J. Krumm. 2011. "Trajectory Processing." In *Computing with Spatial Trajectories*, edited by Y. Zheng and X. Zhou, 1–31. Berlin Heidelberg: Springer.
- Li, Q., Y. Zheng, X. Xie, Y. Chen, W. Liu, and M. Ma. 2008. "Mining User Similarity Based on Location History." In *Proceedings of the 16th Annual ACM International Conference on Advances in Geographic Information Systems*. New York, NY: ACM.
- Ludwig, I., A. Voss, and M. Krause-Traudes. 2011. "A Comparison of the Street Networks of Navteq and OSM in Germany." In *Advancing Geoinformation Science for a Changing World*, edited by S. Geertman, W. Reinhardt and F. Toppen, 65–84. Berlin Heidelberg: Springer.
- Luo, W., H. Tan, L. Chen, and M. N. Lionel. 2013. "Finding Time Period-Based Most Frequent Path in Big Trajectory Data." In *Proceedings of ACM SIGMOD, International Conference of Management of Data*, New York, June 22–27; 713–724.
- Mondzsch, J., and M. Sester. 2011. "Quality Analysis of OpenStreetMap Data Based on Application Needs." *Cartographica: The International Journal for Geographic Information and Geovisualization* 46 (2): 115–125.
- Monreale, A., F. Pinelli, R. Trasarti, and F. Giannotti. 2009. "WhereNext: A Location Predictor on Trajectory Pattern Mining." *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, 637–646. New York, NY: ACM.
- Mooney, P., and P. Corcoran. 2011. "Characteristics of Heavily Edited Objects in OpenStreetMap." *Future Internet* 4 (1): 285–305.
- Mooney, P., P. Corcoran, and A. C. Winstanley. 2010. "Towards Quality Metrics for Openstreetmap." *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '10*, 514–517. New York, NY: ACM.
- Neis, P., D. Zielstra, and A. Zipf. 2013. "Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions." *Future Internet* 5 (2): 282–300.
- Ng, R. T. 1996. "Spatial Data Mining: Discovering Knowledge of Clusters from Maps." In *Proceedings of 1996 ACM-SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, New York, NY: ACM.
- Ng, R. T., and J. Han. 1994. "Efficient and Effective Clustering Methods for Spatial Data Mining." *Proceedings of VLDB 1994, September*, 12–15. Chile, Santiago.
- OpenStreetBugs. n. d. Accessed April 12, 2014. <http://openstreetbugs.schokokeks.org/>
- OSM Inspector. 2012. "OSM Inspector, Geofabrik Tools." Accessed April 12, 2014. <http://tools.geofabrik.de/osmi/>
- OSM Wiki. n. d. "OpenStreetMap Project Wiki." Accessed April 12, 2014. <http://wiki.openstreetmap.org/wiki/>
- OSMOS. n. d. "OpenStreetMap Oversight Search Engine." Accessed April 12, 2014. <http://osmose.openstreetmap.fr>
- Pang, L. X., S. Chawla, W. Liu, and Y. Zheng. 2013. "On Detection of Emerging Anomalous Traffic Patterns Using GPS Data." *Data and Knowledge Engineering* 87: 357–373.
- Pourabdollah, A. 2014. "OSM-GB: Using Open Source Geospatial Tools to Create OSM Web Services for Great Britain." *OSGeo Journal* 13: 41–50.
- Pourabdollah, A., J. Morley, S. Feldman, and M. Jackson. 2013. "Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps' Road Network." *ISPRS International Journal of Geo-Information* 2 (3): 704–728.
- Salk, C. F., T. Sturn, L. See, S. Fritz, and C. Perger. 2015. "Assessing Quality of Volunteer Crowdsourcing Contributions: Lessons from the Cropland Capture Game." *International Journal of Digital Earth* 2015: 1–17.
- Scobbler GmbH. 2012. "MapDust Website." Accessed April 12, 2014. <http://www.mapdust.com>
- Shang, J., Y. Zheng, W. Tong, E. Chang, N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong. 2014. "Discovering Urban Functional Zones Using Latent Activity Trajectories." *IEEE Transaction on Knowledge and Data Engineering* 27 (3): 1041–1347.
- Yavas, G., D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos. 2005. "A Data Mining Approach for Location Prediction in Mobile Environments." *Data & Knowledge Engineering* 54 (2): 121–146.

- Yuan, N. J., Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong. 2015. "Discovering Urban Functional Zones Using Latent Activity Trajectories." *IEEE Transactions on Knowledge and Data Engineering* 27 (3): 712–725.
- Zhang, L., S. Dalyot, and M. Sester. 2013. "Travel-Mode Classification for Optimizing Vehicular Travel Route Planning." *Progress in Location-Based Services*, 277–295. Berlin Heidelberg: Springer.
- Zheng, Y. 2015. "Trajectory Data Mining: An Overview." *ACM Transaction on Intelligent Systems and Technology* 6 (3): 1–41.
- Zielstra, D., and A. Zipf. 2010. OpenStreetMap Data Quality Research in Germany. GIScience 2010. In Sixth international conference on Geographic Information Science, Zurich, Switzerland.